ISSN: 2633-4828 Vol. 5 No.4, December, 2023

International Journal of Applied Engineering & Technology

PREDICTING CHILDHOOD BMI USING MACHINE LEARNING TECHNIQUES - A CASE STUDY IN SAUDI ARABIA

M. Abdollahian[1*], S.H. Alanazi1[2], L. Tafakori[3], K. A. Almulaihan[4], S. M. ALruwili [5] and O.F. Alenazi [6]

[1,2,3]School of Science, RMIT University, La Trobe Street, Melbourne, Australia

[2]Department of Mathematics, College of Sciences, Northern Border University, Arar, Saudi Arabia

[4]Maternity and Children Hospital, Arar, Saudi Arabia

[5,6]North Medical Tower Hospital, Arar, Saudi Arabia

[1]mali.abdollahian@rmit.edu.au

ABSTRACT

Childhood obesity is a significant public health concern globally. Its prevalence has notably increased in Saudi Arabia over the past two decades. Developing models to predict body mass index (BMI) would aid healthcare practitioners in creating effective weight reduction intervention plans. This paper has deployed Random Forest (RF), Decision Tree (DT), and K-Nearest Neighbour (KNN) to predict the BMI at the onset of childhood obesity in Saudi Arabia (S.A.) and to identify the significant factors associated with it. De-identified data from the Arar and Riyadh regions of S.A. were used to develop the prediction models and to compare their performance using multi-prediction accuracy measures. The dataset includes several medical, demographic, and lifestyle variables. The mean and standard deviation for BMI and age were (31.38 1.06) and (10.84 3.12), respectively, with most BMIs falling between 30 and 34 and ages between 5 and 15 years. KNN outperformed other models (accuracy= 0.997, root mean square errors (RMSE) =0.95, mean absolute error (MAE) = 0.73), followed by Random Forest and Decision Tree. The models identified weight at birth, family income and education status, family obesity history, breastfeeding in the first 6 months, lack of physical activities, birth delivery mode, unhealthy food, and age as significant factors in predicting childhood BMI.

Keywords: BMI, children, Saudi Arabia, Decision Tree, Random Forest, and k-Nearest Neighbour.

INTRODUCTION

Obesity, diabetes, and hypertension pose significant public health challenges in Saudi Arabia, with prevalence rates reaching alarming levels according to survey data [1]. Among children, the prevalence of obesity is particularly concerning, with a substantial portion falling into this category, thereby highlighting the urgent need for effective interventions [2]. Sedentary lifestyles among Saudi children further compound this issue,[3]. Given the severe health consequences associated with obesity, such as heart disease and diabetes [4], there is a critical imperative to develop accurate predictive models that can identify obesity and its key performance indicators (KPIs). These models not only assist in identifying risk factors but also play a crucial role in ranking their significance, thus facilitating the development of effective intervention strategies [5].

The study conducted by [6] investigates the connection between hypertension and childhood obesity. The paper highlights the crucial need for early identification and management of hypertension as a preventive measure against long-term health complications. Childhood Obesity is caused by a lack of physical activity, unhealthy eating habits, and genetic factors and can be the cause of different non-communicable diseases [7].

The WHO definitions [8], [9] for obesity and overweight are as follows: Body Mass Index (BMI) is a simple index of weight for height that is commonly used to classify overweight and obesity in adults. It is defined as a person's weight in kilograms divided by the square of his height in meters (kg/m2). For adults, WHO defines overweight and obesity as a BMI greater than or equal to 25 and a BMI greater than or equal to 30.

For children, age needs to be considered when defining overweight and obesity. For children under 5 years of age: Overweight, BMI greater than 2 standard deviations above the WHO Child Growth Standards median; and obesity BMI greater than 3 standard deviations.

For children aged between 5–19 years, BMI-for-age greater than 1 standard deviation above the WHO Growth Reference median (overweight) and greater than 2 standard deviations (obesity).

In 2016, over 650 million adults (>18 years) were obese and 1.3 billion were overweight [9]. These represent 13% and 39% of the global adult population respectively. In 2020, 40 million children under 5 years of age were obese or overweight. Among children and adolescents of 5-19 years, over 340 million were either overweight or obese in 2016. The current status of these numbers could be much higher. Obesity and overweight were restricted to high-income countries. Currently, however, these problems are increasingly noted in urban areas of developing countries, as lifestyles have changed.

Saudi Arabia's dual characteristics as a nation blending attributes of both high-income and developing countries present a unique challenge in combating obesity. A comprehensive review by [10] revealed that Saudi Arabia ranks 15th globally in obesity prevalence, with rates projected to escalate significantly. By 2017, it was estimated that obesity rates would reach 38.2% among men and a staggering 67.5% among women, with an overall prevalence of 52.9%. Projections for 2022 suggest even higher rates, with 41.4% among men, 77.6% among women, and 59.5% overall [10].

Research focused on children underscores the severity of the issue. In the Majmaah region, a primary-school-based survey by [11] found obesity rates at 10.1% and overweight rates at 18.9%, with higher prevalence among females. Similarly, in Al-Ahsa, [12] the rate among male children aged 7 to 15 years was 29.6%. Recent studies by [13] reported a 19.4% obesity rate among children aged 6-18 years, with a notable gender disparity favouring boys. A case-control study by [14]-[17] revealed elevated rates of overweight and obesity compared to WHO standards, highlighting factors such as family history, lack of physical activity, and high consumption of unhealthy food as significant influencers. Various modeling techniques, such as logistic Regression (LR), ANN, support vector machines (SVMs), DT, and KNN can be used to predict Body Mass Index (BMI) [18]-[22].

MOTIVATION AND OBJECTIVE OF THE RESEARCH

In recent years, Saudi Arabia has witnessed a concerning surge in childhood obesity, as evidenced by a study analyzing data from 20,000 cases in the Eastern Province, revealing a staggering 25% prevalence of overweight and obesity among high school students [13]. This alarming trend underscores the imperative to confront childhood obesity in the country to mitigate associated health risks.

The detrimental consequences of childhood obesity extend beyond physical health, encompassing negative impacts on mental health, self-esteem, and social development [23].

Despite the escalating incidence of childhood obesity in Saudi Arabia, research in this domain lags behind that of developed countries, primarily focusing on risk factor identification within localized settings [17]. Notably, there exists a research gap in modeling BMI to predict childhood obesity in the Saudi context.

To address this gap, the proposed study aims to leverage common machine learning techniques, DT, KNN, and RF using local demographic and clinical data from the Arar and Riyadh regions of Saudi Arabia to predict BMI and identify significant key performance indicators associated with overweight or obesity among children aged 3-19 years old. Thereby mitigating issues associated with delayed obesity diagnosis by facilitating early intervention and preventive measures, to reduce the prevalence of childhood obesity and its adverse health outcomes.

Through its comprehensive approach, this study seeks to contribute to global efforts to address childhood obesity while addressing the specific needs of the Saudi Arabian population.

RESEARCH METHODOLOGY

This section outlines the data collection, development, and evaluation of the prediction models using Decision tree (DT), Random Forest (RF), and K-Nearest Neighbours (KNN) methodologies.

A. Data Collection

The de-identified data for this research has been collected from hospitals in the Arar and Riyadh regions of Saudi Arabia between 2011 and 2021. Ethical approval was obtained from the RMIT University Human Research Ethics Committee in Australia and the Research Ethics Committee of the Ministry of Health in Saudi Arabia. The need for informed consent was waived by the ethics committee as this was a retrospective study of medical records. A total of 300 patient records from 2011-2021 have been extracted for children aged between 3-19. A range of medical, demographic, and lifestyle variables was compiled. The variables collected are those recommended by other researchers as being significant factors for childhood obesity and are listed in Table I.

Table I: List of variables used for this study.

NO	*demographic characteristics*	***NO***	*Child characteristics*
1	Gender	**1**	Sleep Hours
2	Residency	**2**	No of times consuming soft Drinks per day
3	consanguineous marriage	**3**	No of times consuming sweets and chocolate per day
4	Income status	**4**	Exercise for 60 minutes per week
5	The education level of the parent	**5**	No of hours watching TV per day
NO	Medical history	**6**	No of times eating fast food per week
1	Age of child at diagnosis	**7**	Eating while watching TV (yes or no)
2	High and weight at diagnosis	**NO**	Obstetric history
3	BMI at diagnosis	**1**	Birth delivery mode
4	Is there in the family anyone with obesity?		natural (SVD) or caesarean section (CS)
	Second-degree relative	**2**	Gestational Age (in weeks)
	First-degree relative Father – mother - siblings	**3**	Weight at birth (Kg)
5	autism		
6	Diabetes		
7	Nutrition history in the first 6 months? (Breastfeeding -Introduction to Cow's milk – Both)		
8	Time mother started to introduce solid food? (< 6 months - > 6 months)		

The number of males and females in the sample by data collection city is shown in Figure I.

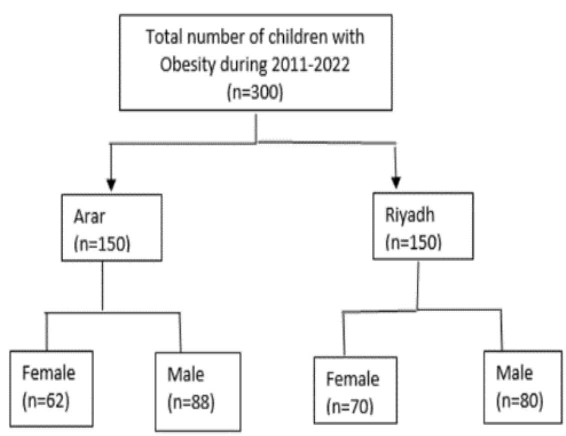

Figure I: Sample size by city and gender.

The data has been analyzed using the statistical techniques of descriptive statistics, and time series, to summarize the data and DT, RF, and KNN to predict BMI and to identify its significant predictors. The assumptions of the various models have also been tested. Further details about the specific models are provided below.

B. Data Analysis and Model Development

1. Descriptive Statistics, T-test and ANOVA

Descriptive statistics including summary statistics and frequency counts have been used to describe the data. The statistics reported include mean, median, standard deviation, minimum and maximum values. Graphs have also been used to investigate the trend of age at the onset of obesity and to extract information on the distribution of age at the onset of obesity. Furthermore, a t-test was employed to assess the disparity in the BMI between different genders. Additionally, the analysis of variance (ANOVA) was utilized to examine the variation in the BMI at the onset of obesity across genders and cities.

2. Methods

Predictive algorithms have been used in a variety of fields, from healthcare to education [24]. The modelling techniques used include DT, RF, and KNN. The dependent variable is the BMI, all variables listed in Table 1 have been used as independent variables. The assumptions of the various models have also been tested. The data was randomly split into a training (70%) and testing set (30%). The best models have been selected based on the smallest Root Mean Squared Error (RMSE), the smallest Mean Absolute Error (MAE), and the largest accuracy of the testing data. Further details about the models are provided below.

a. Decision Tree

DT is a machine-learning technique with a tree-like structure where each internal node represents a test on the input attribute [25], it can develop and present a range of rules for predicting dependent variables using one or more independent variables [26]. A decision tree represents the variables as branches and the target values as leaves. The number of leaves is also known as nodes.

b. Random Forest

RF is an ensemble technique, that is, it fits a range of decision tree models using iterations and recommends the best fit [26]. It can develop and present a range of rules for predicting dependent variables using one or more independent variables. The advancements in RF predictive techniques have been significant in recent years. It has been used to improve predictive accuracy for a variety of applications, including image recognition [27], natural language processing [28], bioinformatics [29], and predicting the onset of type 1 diabetes [30]. Random forests have also been found to outperform traditional algorithms such as support vector machines, boosting, and neural networks [31]. The most recent versions of this algorithm are highly robust and efficient, even in highly complex data sets [32]. This makes it an ideal tool for a wide range of applications, from medical diagnosis to financial forecasting.

c. k-Nearest Neighbour

The KNN model is an example of instance-based learning, which is a form of machine learning where data points are classified based on the values of their closest neighboring points. The KNN algorithm uses the provided data points to find the k-nearest one and then uses that to determine the class of the data point. It is easy to comprehend and can be applied to both classification and regression problems. KNN models are advantageous in health modelling [33] due to their capacity to uncover trends in data that can be overlooked by other modelling approaches [34]. A KNN model can be utilized to ascertain the potential danger of a specific ailment or medical condition dependent on the attributes of a patient's data. As an example, the model may be utilized to identify an individual's likelihood of acquiring diabetes by taking into consideration their age, gender, lifestyle, and additional aspects.

RESULT

A. Descriptive Statistics, T-test and ANOVA

The trend of the reported cases of BMI between 2011 and 2022 (Figure II) illustrates that males have a higher BMI than females over this period. As shown in Table II, the mean BMI in this cohort was 31.38 (SD=1.058078), for males, the mean was 31.30 (SD=1.045066), for females the mean was 31.48 (SD=1.069195). Figure III shows that the BMI is positively skewed and 50% fall between 30 and 31.5 (IQR=1.34). The age is approximately normally distributed, and most children's age falls between 5 to 15 years (Fig III and Table III). The outputs of the t-test (Table IV) show that there are no significant differences in BMI between males and females (p-value=0.13). The results of the ANOVA (Table V) show that there are no significant differences between the main effect of cities, gender, and the interaction between cities and gender on BMI.

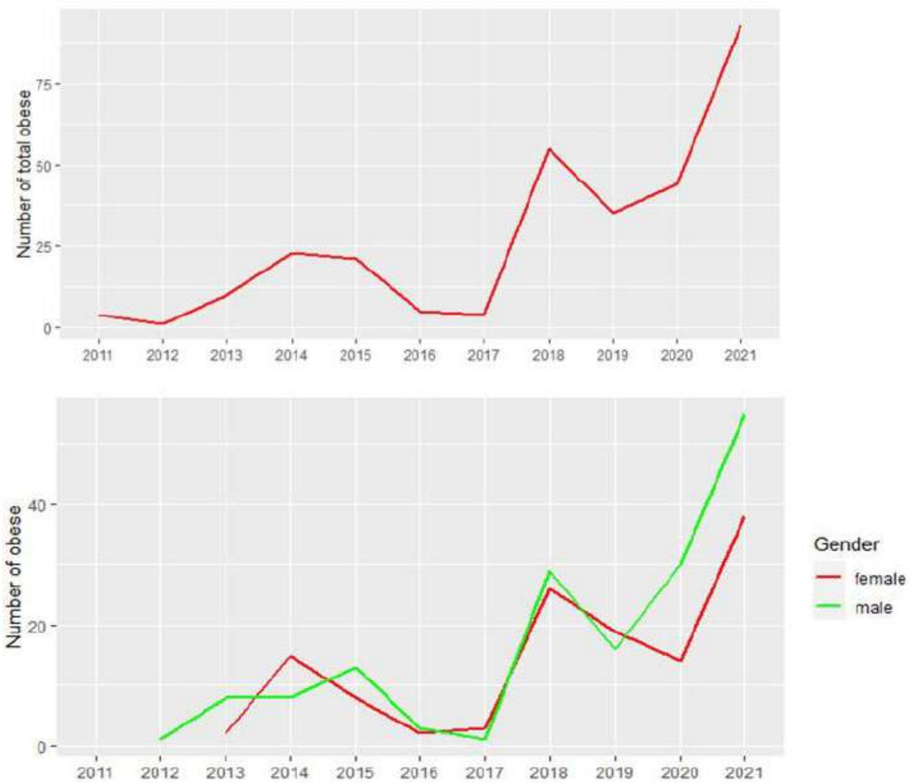

Figure II: Trend of BMI between 2011 and 2022.

Table II. Descriptive statistics for BMI.

	All	*Female*	*Male*
N	300	132	168
Mean	31.38	31.48	31.30
Std. Deviation	1.058078	1.069195	1.045066
Median	31.09	31.16	31.03
Minimum	30.01	30.01	30.01
lower quarter	30.63	30.73	30.47
upper quarter	31.97	32.00	31.83
Maximum	35.58	35.58	34.23

Table III: Descriptive statistics for age.

	All	*Female*	*Male*
N	300	132	168
Mean	10.836	10.890	10.790
Std. Deviation	3.116	3.538	2.7499
Median	10.600	11.500	10.350
Minimum	3.200	3.200	4.300
lower quarter	8.775	8.400	9.200
upper quarter	13.125	13.200	12.430
Maximum	18.600	18.600	18.200

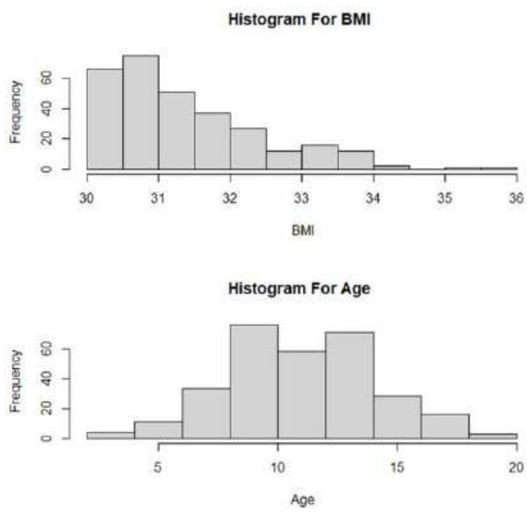

Figure III: BMI and Age distribution.

Table IV: T-test results comparing BMI between males and females.

Gender Level	*N*	*Mean*	*CI*	*t*	*df*	*p-value*
Male	132	31.48409	(-0.055151 0.429623)	1.5206	278.42	0.1295
Female	168	31.29685				

Table V: ANOVA results comparing BMI between gender and cities.

Variables		*Mean Difference*	*CI*	*P-value*
Gender	Male-Female	-0.18724	(-0.429318 0.0548464)	0.1290
Cities	Riyadh-Arar	0.11303	(-0.127469 0.353522)	0.3558
Gender: City	Male: Arar- Female: Arar	-0.11986	(-0.572934 0.333205)	0.9033
	Female: Riyadh -Female: Arar	0.17168	(-0.304854 0.6482241)	0.7883

B. Modelling Results

This section presents the modeling results of the Decision Tree, Random Forest, and k-Nearest Neighbour.

1. Decision Tree

The hyperparameter governing the randomization of the split variable feature in DT which is commonly denoted as the "cp" parameter has been considered to determine the tree's complexity. Through 10-fold cross-validation based on the training data, we observed that the best (cp) for the DT is 0.1.

The model validation statistics for the training and testing data presented in Table VI show RMSE=1.00, MAE=0.84, and accuracy=0.992.

2. Random Forest

The hyperparameter that controls the split variable randomization feature of RF is often referred to mtry. This is the number of variables randomly sampled as candidates at each split, and it helps to balance the trade-off between a low correlation and reasonable strength. Through 10-fold cross-validation based on the training data, we observed that the best mtry for the RF is 2. The model validation statistics for the testing and training data presented in Table VI show RMSE=1.102, MAE=0.854, and accuracy=0.994.

3. K-Nearest Neighbour

The hyperparameter that controls the split variable randomization feature of KNN is often referred to (K)parameter which controls the number of variables in the model. Through 10-fold cross-validation based on the training data, we observed that the best (K) for the KNN is 2. The model validation statistics for the testing data presented in Table VI show RMSE=0.947, MAE=0.735, and accuracy=0.997.

C. Model Validation

Root Mean Square Error (RMSE) is a measure of model performance or prediction error and is calculated by taking the square root of the mean of the squared differences between the predicted values and the actual values [35]. It is a more robust measure of model performance than Mean Absolute Error (MAE) because it penalizes large errors. Mean Absolute Error (MAE) is a measure of model performance or prediction error and is calculated by taking the mean of the absolute differences between the predicted values and the actual values. It is a less robust measure of model performance than Root Mean Square Error (RMSE) because it does not penalize large errors as much [35]. Validation of the models was carried out using the maximum accuracy and minimum RMSE and MAE. The model validation statistics for all the models are shown below. The results presented in Table VI show that KNN outperformed other models followed by RF and DT.

Table VI: Model validation statistics

Models	*Training data*			*Testing data*		
	RMSE	MAE	accuracy	RMSE	MAE	Accuracy
RF	0.9494	0.7566	0.9989	1.1019	0.85436	0.9939
DT	1.0724	0.85936	0.9990	1.0001	0.8398	0.9922
KNN	1.0145	0.8079	0.9978	0.9467	0.73506	0.9967

The relative importance of each independent variable was determined, and the significant variables identified by each models are listed in Table VII.

Table VII: Ranked significant variables identified by each model.

Model	***Significant variables***
Random Forest (RF).	autism yes Child sleep hours regular Gestational Age weeks Weight at birth Kg soft drinks 3 or more times Education level of Father University Income status middle Consanguineous marriage yes Eat while watching TV yes Exercise for 60 minutes or more 1-2 time sweet chocolate per day More than twice Nutrition in first 6 months Introduction to Cow's milk
Decision Tree.	sweet chocolate per day More than twice First degree relative yes Father Yes.

Model	*Significant variables*
	soft drinks 3 or more times Fast food 3 or more times Watch TV per day 3-4 hours Nutrition in first 6 months Breastfeeding Gestational Age weeks Weight at birth Kg Gender female Education level of mother University
KNN	First degree relative yes Father Yes sweet chocolate per day More than twice Fast food 3 or more times soft drinks 3 or more times autism yes Nutrition in first 6 months Introduction to Cow's mil. Solid food more than 6 months Birth delivery mode Caesarean section (CS) Watch TV per day 3-4 hours Gender female Income status Middle

CONCLUSION

Obesity has become a widespread issue among both adults and children worldwide, and Saudi Arabia is no exception. The country faces a particular challenge in reducing childhood obesity, as most children lead a sedentary lifestyle.

The present study employed local de-identified data for children aged 3-19 years from the Arar and Riyadh regions of Saudi Arabia to predict childhood BMI and to identify the significant factors contributing to childhood BMI in Saudi Arabia. The mean BMI was 31.38 (SD=1.058) with no significant difference between males and females or cities. The most common age was between 5 to 15 years.

Decision Tree, Random Forest, and K-Nearest Neighbour were deployed to predict BMI. The results based on RMSE, MAE, and accuracy show that KNN (RMSE=0.947, MAE=0.735, and accuracy=0.997) outperformed RF (RMSE=1.102, MAE =0.854, and accuracy=0.994) and DT (RMSE=1.00, MAE=0.84 and accuracy=0.992). The models identified weight at birth, family income and education status, family obesity history, breastfeeding in the first 6 months, lack of physical activities, birth delivery mode, unhealthy food, and age as significant factors in predicting childhood BMI.

Some of these factors were identified as risk factors for obesity and overweight by other international researchers [13],[22],[36],[37], Specifically, education level, age, and glucose level were found to be risk factors for obesity and overweight in their study.

These findings are also consistent with other Suadi studies [12],[14] which found gender, age, family history of obesity, family income, and physical activity are associated with an increased risk of childhood overweight and obesity. Our results presented here support the suggestion put forth by [14] regarding the requirement for tailored preventive and interventional measures to mitigate the high prevalence of childhood overweight and obesity in Saudi Arabia. Based on the significant variables identified in this research, some of the recommendations for reducing childhood obesity are: increase access to healthy food options by providing more fresh fruit and vegetable markets in low-income communities; encourage physical activity in children by providing more safe and accessible outdoor spaces, such as playgrounds and parks; implement nutrition education programs in schools to teach children about healthy eating habits; create public awareness campaigns to educate parents and families about the importance of a healthy lifestyle; restrict marketing of unhealthy foods and beverages to children; encourage breastfeeding; promote and support breastfeeding-friendly workplaces; and Offer free or low-cost nutrition counseling services to families.

The best model identified through this research can be used to predict children's BMI and identify children who might be at high risk of obesity. This information can then be utilised to mitigate the risks in such populations.

Some possible limitations of this study, which are lack of comprehensive data on a variety of health-related factors may limit the accuracy of predictions. Also, the findings of the study may not apply to other populations that have different dietary habits and lifestyles.

REFERENCES

[1] Mokdad, A. (2014, March 10). Saudi Health Interview Survey finds high rates of chronic diseases in the Kingdom of Saudi Arabia. Retrieved May 3, 2019, from http://www.healthdata.org/news-release/saudi-health-interview-survey-finds-high-rates-chronic-diseases-kingdom-saudi-arabia.

[2] Centers for Disease Control and Prevention. (2016). Childhood obesity causes & consequences. Division of Nutrition, Physical Activity, and Obesity. Retrieved Mar28, 2021from http://www.cdc.gov/obesity/childhood/causes.html.

[3] World Health Organization. (2012). Country cooperation strategy for WHO and Saudi Arabia 2012-2016 [Brochure]. World Health Organization. Regional Office for the Eastern Mediterranean. Retrieved March 28, 2021, from http://www.who.int/countryfocus/cooperation_strategy/ccs_sau_en.pdf.

[4] Jin, J. (2013). Obesity and the Heart. Jama, 310(19), 2113-2113.

[5] Colmenarejo, G. (2020). Machine learning models to predict childhood and adolescent obesity: a review. Nutrients, 12(8), 2466.

[6] Wuehl, E.: Hypertension in childhood obesity. Acta Paediatrica 108(1), 37–43 (2019)

[7] WHO. (2018, February 16). Obesity and overweight. Retrieved May 3, 2019, from World Health Organisation: https://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight.

[8] Kelishadi, R., Heidari-Beni, M.: Prevention and control of childhood obesity: the backbone in prevention of noncommunicable disease. Primordial Prevention of Non Communicable Disease, 61–66 (2019)

[9] WHO. (2020, March 3). Obesity and Overweight. Retrieved April 19, 2020, from WHO: https://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight

[10] Alqarni, S. M. (2016). A review of prevalence of obesity in Saudi Arabia. Journal of Obesity & Eating Disorders, 2(2), 25. doi:10.21767/2471-8203.100025

[11] Abdalla, S. M., Alsaif, B. A., Al Jasser, S. J., & Al Sultan, A. S. (2017). Prevalence of Obesity and Overweight among Primary School Children in Majmaah, Saudi Arabia. Majmaah Journal of Health Sciences, 5(1), 30-40. doi:10.5455/mjhs.2017.01.005

[12] Saleh, A. A., Alhaiz, A. S., Khan, A. R., Al-Quwaidhi, A. J., Aljasim, M., Almubarak, A., ... & AlYateem, A. (2017). Prevalence of obesity in school children and its relation to lifestyle behaviors in Al-Ahsa district of Saudi Arabia. *Glob J Health Sci, 9*(12), 1-80.

[13] Albaker, W., Saklawi, R., Bah, S., Motawei, K., Futa, B., & Al-Hariri, M. (2022). What is the current status of childhood obesity in Saudi Arabia?: Evidence from 20,000 cases in the Eastern Province: a cross-sectional study. Medicine, 101(27).

[14] Aljassim, H., & Jradi, H. (2021). Childhood overweight and obesity among the Saudi population: a case-control study among school children. Journal of Health, Population and Nutrition, 40(1), 1-9.

[15] Al-Hazzaa, H.M., Alrasheedi, A.A., Alsulaimani, R.A., Jabri, L., Alhowikan, A.M., Alhussain, M.H., Bawaked,R.A., Alqahtani, S.A.: Prevalence of overweight and obesity among saudi children: A comparison of two widelyused international standards and the national growth references. Frontiers in Endocrinology 13 (2022)

[16] AlEnazi, S., AlAjlan, R., AlKhalaf, H., Abolfotouh, M., Alharbi, O., Alfawaz, R., Aldebasi, B., Alghnam, S.:Prevalence of obesity among children and adolescents in saudi arabia: A multicenter population-based study.Saudi Journal of Medicine & Medical Sciences 11(1), 19 (2023)

[17] Shaikh, M. A., Al Sharaf, F., Shehzad, K., Shoukat, F., Naeem, Z., Al Harbi, S., . . . Al Motairi, S. (2016). Prevalence and trends of overweight and obesity amongst Saudi school children, a study done by using three noninvasive methods. International journal of health sciences, 10(3), 381-387. Retrieved April 19, 2020, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5003581/

[18] Chiong, R., Fan, Z., Hu, Z., & Chiong, F. (2021). Using an improved relative error support vector machine for body fat prediction. Computer Methods and Programs in Biomedicine, 198, 105749.

[19] Garcia-Carretero, R., Vigil-Medina, L., Mora-Jimenez, I., Soguero-Ruiz, C., Barquero-Perez, O., & Ramos-Lopez, J. (2020). Use of a K-nearest neighbors model to predict the development of type 2 diabetes within 2 years in an obese, hypertensive population. Medical & biological engineering & computing, 58, 991-1002.

[20] Kuczmarski, R. J., Flegal, K. M., Campbell, S. M., & Johnson, C. L. (1994). Increasing prevalence of overweight among US adults. The Journal of the American Medical Association, 272(3), 205-211.

[21] Baltasar, A., Perez, N., Serra, C., Bou, R., Bengochea, M., & Borrás, F. (2011). Weight loss reporting: predicted body mass index after bariatric surgery. Obesity surgery, 21(3), 367-372.

[22] Meng, X. H., Huang, Y. X., Rao, D. P., Zhang, Q., & Liu, Q. (2013). Comparison of three data mining models for predicting diabetes or prediabetes by risk factors. The Kaohsiung journal of medical sciences, 29(2), 93-99.

[23] Sahoo, K., Sahoo, B., Choudhury, A. K., Sofi, N. Y., Kumar, R., & Bhadoria, A. S. (2015). Childhood obesity: Causes and consequences. Journal of Family Medicine and Primary Care, 4(2), 187–192. https://doi.org/10.4103/2249-4863.154628

[24] A Lantz, B. (2019). Machine learning with R: expert techniques for predictive modeling. Packt publishing ltd.

[25] Uddin, Shahadat, Stephen Ong, and Haohui Lu. "Machine learning in project analytics: A data-driven framework and case study." Scientific Reports 12.1 (2022): 15252

[26] Bonaccorso, G. (2017). Machine learning algorithms. Packt Publishing Ltd.

[27] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In Advances in neural information processing systems (pp. 1097-1105).

[28] Collobert, R., Weston, J., Bottou, L., Karlen, M., Kavukcuoglu, K., & Kuksa, P. (2011). Natural language processing (almost) from scratch. Journal of Machine Learning Research, 12, 2493-2537.

[29] Kursa, M.B., Rudnicki, W.R.: Feature selection with the boruta package. Journal of statistical software 36,1–13 (2010)

[30] Alazwari, A., Abdollahian, M., Tafakori, L., Johnstone, A., Alshumrani, R. A., Alhelal, M. T., ... & Alkhaldi, A. R. (2022). Predicting age at onset of type 1 diabetes in children using regression, artificial neural network and Random Forest: A case study in Saudi Arabia. Plos one, 17(2), e0264118.

[31] Chang, C.-C., & Lin, C.-J. (2011). LIBSVM: A library for support vector machines. ACM Transactions on Intelligent Systems and Technology, 2(3), 27.

[32] Bansal, M., Kumar, M., Sachdeva, M., & Mittal, A. (2021). Transfer learning for image classification using VGG19: Caltech-101 image data set. Journal of ambient intelligence and humanized computing, 1-12.

[33] Almutairi, E. S., & Abbod, M. F. (2023). Machine Learning Methods for Diabetes Prevalence Classification in Saudi Arabia. Modelling, 4(1), 37-55.

[34] Xing, W., & Bei, Y. (2019). Medical health big data classification based on KNN classification algorithm. IEEE Access, 8, 28808-28819.

[35] Willmott, Cort J., and Kenji Matsuura. "Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance." Climate research 30.1 (2005): 79-82.

[36] Sahoo, K., Sahoo, B., Choudhury, A. K., Sofi, N. Y., Kumar, R., & Bhadoria, A. S. (2015). Childhood obesity: Causes and consequences. Journal of Family Medicine and Primary Care, 4(2), 187–192. https://doi.org/10.4103/2249-4863.154628

[37] Ferenci, T., & Kovacs, L. (2018). Predicting body fat percentage from anthropometric and laboratory measurements using artificial neural networks. Applied Soft Computing, 67, 834-839.